

Feature Matrix & Product Roadmap

Patrol Sight

HumanityAI'd — July 2026

·CONFIDENTIAL

1. Feature matrix (v1.1.7 — running today)

Capability	Status	Evidence / notes
Hands-free face recognition in the officer's lens	Live	~68 ms typical glasses-to-label, measured stage by stage
Hardware H.264 streaming from AR glasses	Live	Camera2 → hardware encoder → RTP/UDP; 2560×1440 at 20 Mbps, 0% packet loss
SCRFD detection + ArcFace 512-d embeddings on TensorRT FP16	Live	Model choice benchmarked across 0-90° yaw
GPU FAISS IVF-PQ watchlist search	Live	93,728 persons in production; P95 8.5 ms over 1,931 searches
Two-layer tracking (spatial + identity cache)	Live	Search skipped on 97.6% of frames; GPU utilisation 85% → 40%
Colour-coded category labels in-lens	Live	Encoded name @@category@@ #colour format, 10 colours verified
Label stability (smoothing, stale eviction, duplicate suppression)	Live	EMA smoothing + spatial matching + IoU coasted-track suppression
Long-range mode (8-10 m)	Live	Raised detector resolution, reduced frame rate
Bulk watchlist enrolment	Live	40-60 images/second
Live index rebuild + hot reload	Live	No restart, no retraining
Face-quality assessment	Live	Suppresses unusable frames
Control console — live device monitor	Live	Per-device fps, latency, face counts, annotated preview
Targets, labels, tracking history, device tracking	Live	C2 console

Capability	Status	Evidence / notes
Alert publication on internal message bus	Live	For downstream command-and-control integration
Role-based operator accounts + audit trail	Live	Audit middleware + operator audit-log view
Privacy mode — send faceprint, not image	Live	512 floats (~2 KB) vs image (~100 KB)
Per-device sessions, one-command provisioning	Live	Device-keyed sessions; <code>deploy-glasses.sh</code>
On-prem single-host deployment, 34-service stack	Live	One-command deploy/health/update/teardown
Boot recovery incl. GPU driver race handling	Live	systemd unit + Swarm state + GPU startup wrapper
Zero-trust internal messaging + encrypted console payloads	Live	Per-service bus authentication; AES-256-GCM
Backpressure-safe ingest	Live	Newest-frame-wins + frame-rate-capped worker
Browser-webcam test mode	Live	Validate recognition without glasses

2. Known limitations (honest)

- **Recognition envelope:** reliable to ~5 m, maximum 6–7 m in good lighting; faces below 40×40 pixels are rejected rather than guessed. Long-range mode reaches 8–10 m at reduced frame rate.
- **One GPU server ≈ one site.** The multi-GPU configuration is written into the stack file but not yet enabled and measured.
- **Concurrent-device capacity per GPU is not yet published** as a measured number — it is bounded by decode and detection, not by search. We measure it during the pilot and contract on that figure.
- **The glasses transport is plaintext UDP**, safe only on an isolated VLAN with no uplink. Encrypted transport is Phase 2 hardening.
- **Credential and network-surface hardening is incomplete** — default credentials in the stack file, a committed API key in the app source, and datastore ports published on the host. Phase 0 of the hardening plan, and disclosed in full in the Technical Due-Diligence Report.
- **No third-party watchlist connector.** The watchlist is the customer's own, enrolled through our console; national or interoperable database sync is roadmap v1.5.
- **Model licensing is unresolved** for commercial deployment — the recognition weights derive from a research-licensed lineage. This is the top-priority pre-revenue item.
- **No alert back to the wearer other than the visual label** — no audio or haptic channel yet.
- **No automated test suite or CI** at stack and edge-app level.

3. Roadmap

v1.2 — Deployment-ready hardening (0–3 months)

- Rotate all credentials; per-device provisioned tokens in device keystore; secrets out of source and out of the app package
- Encrypted glasses transport (SRTP/DTLS or VLAN-level VPN); collapse non-nginx host ports to the private LAN IP
- Fail loudly instead of silently falling back when the TensorRT engine cannot load
- Remove the Docker-socket mount from the operations container; RBAC across console and engine APIs
- Automated tests + CI; deployed-stack smoke tests; connection-pooler transaction mode
- **Resolve model licensing** (commercial licence or replacement weights)

v1.5 — Operational reach (3-6 months)

- **Audio and haptic alerts to the wearer** for high-confidence matches — eyes-up, ears-in
- **External watchlist connectors** — sync identities from the customer's existing systems or authorised third-party lists
- **Command-and-control / VMS integration pack** — publish alerts into the customer's existing operations platform rather than beside it
- Multi-GPU enabled and measured; published concurrent-device ceiling per GPU class
- Fleet management: device health, battery and thermal telemetry, over-the-air app update
- Thermal-adaptive capture ladder for long shifts

v2.0 — Beyond identification (6-12 months)

- **Behaviour and anomaly analytics** (loitering, crowd surge, restricted-area entry)
- Multi-camera fusion: correlate the officer's view with fixed cameras for continuous tracking across a site
- Vehicle and licence-plate recognition from the same wearable stream
- Offline degraded mode: on-device matching against a cached high-priority subset when the link drops
- Live operations dashboard with attention and coverage heat-maps

v2.x — Platform

- Multi-site command console for ministries and estate groups
- Federated watchlists with per-site scoping and cross-site audit
- Evidential export packages formatted for judicial review
- Accessibility and localisation of the officer and operator interfaces (Arabic-first console)

4. Differentiators to protect

1. **Measured sub-100 ms glasses-to-label** — published stage by stage, not asserted. Nobody in this category publishes a latency budget.
2. **The two-layer tracker** — eliding 97.6% of vector searches is what makes continuous patrol recognition affordable on one GPU. This is the core IP.
3. **On-premises by construction** — recognition, watchlist and audit never leave the customer's network, which is where cloud recognition vendors cannot follow under GCC data-protection law.
4. **Customer-owned watchlist, no scraped data, human verification required** — the governance posture is a competitive asset, not a compliance cost.
5. **A real operational envelope** — one-command deploy, boot recovery, external volumes, release snapshots, zero-trust internal messaging. This is deployed infrastructure, not a demo.